

# NONPARAMETRIC ESTIMATE OF SPECTRAL DENSITY FUNCTIONS OF SAMPLE COVARIANCE MATRICES: A FIRST STEP

BY BING-YI JING<sup>1</sup>, GUANGMING PAN<sup>2</sup>, QI-MAN SHAO<sup>3</sup>  
 AND WANG ZHOU<sup>4</sup>

*Hong Kong University of Science and Technology, Nanyang Technological  
 University, Hong Kong University of Science and Technology  
 and National University of Singapore*

The density function of the limiting spectral distribution of general sample covariance matrices is usually unknown. We propose to use kernel estimators which are proved to be consistent. A simulation study is also conducted to show the performance of the estimators.

**1. Introduction.** Suppose that  $X_{ij}$  are independent and identically distributed (i.i.d.) real random variables. Let  $\mathbf{X}_n = (X_{ij})_{p \times n}$  and  $\mathbf{T}_n$  be a  $p \times p$  nonrandom Hermitian nonnegative definite matrix. Consider the random matrices

$$\mathbf{A}_n = \frac{1}{n} \mathbf{T}_n^{1/2} \mathbf{X}_n \mathbf{X}_n^T \mathbf{T}_n^{1/2}.$$

When  $EX_{11} = 0$  and  $EX_{11}^2 = 1$ ,  $\mathbf{A}_n$  can be viewed as a sample covariance matrix drawn from the population with covariance matrix  $\mathbf{T}_n$ . Moreover, if  $\mathbf{T}_n$  is another sample covariance matrix, independent of  $\mathbf{X}_n$ , then  $\mathbf{A}_n$  is a Wishart matrix.

Sample covariance matrices are of paramount importance in multivariate analysis. For example, in principal component analysis, we need to estimate

---

Received January 2010; revised May 2010.

<sup>1</sup>Supported in part by Hong Kong RGC Grants HKUST6011/07P and HKUST6015/08P.

<sup>2</sup>Supported in part by Grant M58110052 at the Nanyang Technological University.

<sup>3</sup>Supported in part by Hong Kong RGC CERG 602608.

<sup>4</sup>Supported in part by Grant R-155-000-083-112 at the National University of Singapore.

*AMS 2000 subject classifications.* Primary 15A52, 60F15, 62E20; secondary 60F17.

*Key words and phrases.* Sample covariance matrices, Stieltjes transform, nonparametric estimate.

This is an electronic reprint of the original article published by the  
 Institute of Mathematical Statistics in *The Annals of Statistics*,  
 2010, Vol. 38, No. 6, 3724–3750. This reprint differs from the original in  
 pagination and typographic detail.

eigenvalues of sample covariance matrices in order to obtain an interpretable low-dimensional data representation. The matrices consisting of contemporary data are usually large, with the number of variables proportional to the sample size. In this setting, fruitful results have accumulated since the celebrated Marcenko and Pastur law [8] was discovered; see the latest monograph of Bai and Silverstein [4] for more details.

The basic limit theorem regarding  $\mathbf{A}_n$  concerns its empirical spectral distribution  $F^{\mathbf{A}_n}$ . Here, for any matrix  $\mathbf{A}$  with real eigenvalues, the empirical spectral distribution  $F^{\mathbf{A}}$  is given by

$$F^{\mathbf{A}}(x) = \frac{1}{p} \sum_{k=1}^p I(\lambda_k \leq x),$$

where  $\lambda_k$ ,  $k = 1, \dots, p$ , denote the eigenvalues of  $\mathbf{A}$ .

Suppose the ratio of the dimension to the sample size  $c_n = p/n$  tends to  $c$  as  $n \rightarrow \infty$ . When  $\mathbf{T}_n$  becomes the identity matrix,  $F^{\mathbf{A}_n}$  tends to the so-called Marcenko and Pastur law with the density function

$$f_c(x) = \begin{cases} (2\pi cx)^{-1} \sqrt{(b-x)(x-a)}, & a \leq x \leq b, \\ 0, & \text{otherwise.} \end{cases}$$

It has point mass  $1 - c^{-1}$  at the origin if  $c > 1$ , where  $a = (1 - \sqrt{c})^2$  and  $b = (1 + \sqrt{c})^2$  (see Bai and Silverstein [4]).

In the literature, it is also common to study

$$\mathbf{B}_n = \frac{1}{n} \mathbf{X}_n^T \mathbf{T}_n \mathbf{X}_n$$

since the eigenvalues of  $\mathbf{A}_n$  and  $\mathbf{B}_n$  differ by  $|n - p|$  zero eigenvalues. Thus,

$$(1.1) \quad F^{\mathbf{B}_n}(x) = \left(1 - \frac{p}{n}\right) I(x \in [0, \infty)) + \frac{p}{n} F^{\mathbf{A}_n}(x).$$

When  $F^{\mathbf{T}_n}$  converges weakly to a nonrandom distribution  $H$ , Marcenko and Pastur [8], Yin [16] and Silverstein [13] proved that, with probability one,  $F^{\mathbf{B}_n}(x)$  converges in distribution to a nonrandom distribution function  $\underline{F}_{c,H}(x)$  whose Stieltjes transform  $\underline{m}(z) = m_{\underline{F}_{c,H}}(z)$  is, for each  $z \in \mathcal{C}^+ = \{z \in \mathcal{C} : \Im z > 0\}$ , the unique solution to the equation

$$(1.2) \quad \underline{m} = - \left( z - c \int \frac{t dH(t)}{1 + t \underline{m}} \right)^{-1}.$$

Here, the Stieltjes transform  $m_F(z)$  for any probability distribution function  $F(x)$  is defined by

$$(1.3) \quad m_F(z) = \int \frac{1}{x - z} dF(x), \quad z \in \mathcal{C}^+.$$

Therefore, from (1.1), we have

$$(1.4) \quad \underline{F}_{c,H}(x) = (1 - c)I(x \in [0, \infty)) + cF_{c,H}(x),$$

where  $F_{c,H}(x)$  is the limit of  $F^{\mathbf{A}_n}(x)$ . As a consequence of this fact, we have

$$(1.5) \quad \underline{m}(z) = -\frac{1 - c}{z} + cm(z).$$

Moreover,  $\underline{m}(z)$  has an inverse,

$$(1.6) \quad z(\underline{m}) = -\frac{1}{\underline{m}} + c_n \int \frac{t dH(t)}{1 + t\underline{m}}.$$

Relying on this inverse, Silverstein and Choi [14] carried out a remarkable analysis of the analytic behavior of  $\underline{F}_{c,H}(x)$ .

When  $\mathbf{T}_n$  becomes the identity matrix, there is an explicit solution to (1.2). In this case, from (1.1), we see that the density function of  $\underline{F}_{c,H}(x)$  is

$$\underline{f}_{c,I}(x) = (1 - c)I(c < 1)\delta_0 + cf_c(x),$$

where  $\delta_0$  is the point mass at 0. Unfortunately, there is no explicit solution to (1.2) for general  $\mathbf{T}_n$ . Although we can use  $F^{\mathbf{A}_n}(x)$  to estimate  $F_{c,H}(x)$ , we cannot make any statistical inference on  $F_{c,H}(x)$  because there is, as far as we know, no central limit theorem concerning  $(F^{\mathbf{A}_n}(x) - F_{c,H}(x))$ . Actually, it is argued in Bai and Silverstein [4] that the process  $n(F^{\mathbf{A}_n}(x) - F_{c,H}(x))$ ,  $x \in (-\infty, \infty)$ , does not converge to a nontrivial process in any metric space. This makes us want to pursue other ways of understanding the limiting spectral distribution  $F_{c,H}(x)$ .

This paper is part of a program to estimate the density function  $f_{c,H}(x)$  of the limiting spectral distribution  $F_{c,H}(x)$  of sample covariance matrices  $\mathbf{A}_n$  by kernel estimators. In this paper, we will prove the consistency of those estimators as a first step.

**2. Methodology and main results.** Suppose that the observations  $X_1, \dots, X_n$  are i.i.d. random variables with an unknown density function  $f(x)$  and  $F_n(x)$  is the empirical distribution function determined by the sample. A popular nonparametric estimate of  $f(x)$  is then

$$(2.1) \quad \hat{f}_n(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x - X_j}{h}\right) = \frac{1}{h} \int K\left(\frac{x - y}{h}\right) dF_n(y),$$

where the function  $K(y)$  is a Borel function and  $h = h(n)$  is the bandwidth which tends to 0 as  $n \rightarrow \infty$ . Obviously,  $\hat{f}_n(x)$  is again a probability density function and, moreover, it inherits some smooth properties of  $K(x)$ , provided the kernel is taken as a probability density function. Under some regularity conditions on the kernel, it is well known that  $\hat{f}_n(x) \rightarrow f(x)$  in

some sense (with probability one, or in probability). There is a huge body of literature regarding this kind of estimate. For example, one may refer to Rosenblatt [10], Parzen [9], Hall [7] or the book by Silverman [12].

Informed by (2.1), we propose the following estimator  $f_n(x)$  of  $f_{c,H}(x)$ :

$$(2.2) \quad f_n(x) = \frac{1}{ph} \sum_{i=1}^p K\left(\frac{x - \mu_i}{h}\right) = \frac{1}{h} \int K\left(\frac{x - y}{h}\right) dF^{\mathbf{A}_n}(y),$$

where  $\mu_i$ ,  $i = 1, \dots, p$ , are eigenvalues of  $\mathbf{A}_n$ . It turns out that  $f_n(x)$  is a consistent estimator of  $f_{c,H}(x)$  under some regularity conditions.

Suppose that the kernel function  $K(x)$  satisfies

$$(2.3) \quad \sup_{-\infty < x < \infty} |K(x)| < \infty, \quad \lim_{|x| \rightarrow \infty} |xK(x)| = 0$$

and

$$(2.4) \quad \int K(x) dx = 1, \quad \int |K'(x)| dx < \infty.$$

**THEOREM 1.** *Suppose that  $K(x)$  satisfies (2.3) and (2.4). Let  $h = h(n)$  be a sequence of positive constants satisfying*

$$(2.5) \quad \lim_{n \rightarrow \infty} nh^{5/2} = \infty, \quad \lim_{n \rightarrow \infty} h = 0.$$

*Moreover, suppose that all  $X_{ij}$  are i.i.d. with  $EX_{11} = 0$ ,  $\text{Var}(X_{11}) = 1$  and  $EX_{11}^{16} < \infty$ . Also, assume that  $c_n \rightarrow c \in (0, 1)$ . Let  $\mathbf{T}_n$  be a  $p \times p$  nonrandom symmetric positive definite matrix with spectral norm bounded above by a positive constant such that  $H_n = F^{\mathbf{T}_n}$  converges weakly to a nonrandom distribution  $H$ . In addition, suppose that  $F_{c,H}(x)$  has a compact support  $[a, b]$  with  $a > 0$ . Then,*

$$f_n(x) \longrightarrow f_{c,H}(x) \quad \text{in probability uniformly in } x \in [a, b].$$

**REMARK 1.** We conjecture that the condition  $EX_{11}^{16}$  can be reduced to  $EX_{11}^4 < \infty$ .

When  $\mathbf{T}_n$  is the identity matrix, we have a slightly better result.

**THEOREM 2.** *Suppose that  $K(x)$  satisfies (2.3) and (2.4). Let  $h = h(n)$  be a sequence of positive constants satisfying*

$$(2.6) \quad \lim_{n \rightarrow \infty} nh^2 = \infty, \quad \lim_{n \rightarrow \infty} h = 0.$$

*Moreover, suppose that all  $X_{ij}$  are i.i.d. with  $EX_{11} = 0$ ,  $\text{Var}(X_{11}) = 1$  and  $EX_{11}^{12} < \infty$ . Also, assume that  $c_n \rightarrow c \in (0, 1)$ . Denote the support of the MP law by  $[a, b]$ . Let  $\mathbf{T}_n = \mathbf{I}$ . Then,*

$$\sup_{x \in [a, b]} |f_n(x) - f_c(x)| \longrightarrow 0 \quad \text{in probability.}$$

Theorem 1 also gives the estimate of  $F_{c,H}(x)$ , as below.

COROLLARY 1. *Under the assumptions of Theorem 1, correspondingly,*

$$(2.7) \quad F_n(x) \rightarrow F_{c,H}(x) \quad \text{in probability,}$$

where

$$(2.8) \quad F_n(x) = \int_{-\infty}^x f_n(t) dt.$$

Corollary 1 and the Helly–Bray lemma ensure that we have the following.

COROLLARY 2. *Under the assumptions of Theorem 1, if  $g(x)$  is a continuous bounded function, then*

$$(2.9) \quad \int g(x) dF_n(x) \rightarrow \int g(x) dF_{c,H}(x) \quad \text{in probability.}$$

In order to prove consistency of the nonparametric estimates, we need to develop a convergence rate for  $F^{\mathbf{A}_n}$ . When  $\mathbf{T}_n = \mathbf{I}$ , Bai [1] developed a Berry–Esseen-type inequality and investigated the convergence rate of  $EF^{\mathbf{A}_n}$ . Later, Götze and Tikhomirov [6] improved the Berry–Esseen-type inequality and obtained a better convergence rate. For general  $\mathbf{T}_n$ , we establish the following convergence rate.

THEOREM 3. *Under the assumptions of Theorem 1,*

$$(2.10) \quad \sup_x |EF^{\mathbf{A}_n}(x) - F_{c_n,H_n}(x)| = O\left(\frac{1}{n^{2/5}}\right)$$

and

$$(2.11) \quad E \sup_x |F^{\mathbf{A}_n}(x) - F_{c_n,H_n}(x)| = O\left(\frac{1}{n^{2/5}}\right).$$

REMARK 2. Under the fourth moment condition, that is,  $EX_{11}^4 < \infty$ , we conjecture that the above rate  $O(n^{-2/5})$  could be improved to  $O(n^{-1}\sqrt{\log n})$ .

**3. Applications.** Let us demonstrate some applications of Theorems 1, 2 and their corollaries. Since  $F_{c,H}(x)$  does not have an explicit expression (except for some special cases), we may now use  $F_n(x)$  to estimate it, by Corollary 1. More importantly,  $F_n(x)$  has some smoothness properties, which  $F^{\mathbf{A}_n}$  does not have.

We first consider an example in wireless communication. Consider a synchronous CDMA system with  $n$  users and processing gain  $p$ . The discrete-time model for the received signal  $\mathbf{Y}$  is given by

$$(3.1) \quad \mathbf{Y} = \sum_{k=1}^n x_k \mathbf{h}_k + \mathbf{W},$$

where  $x_i \in \mathcal{R}$  and  $\mathbf{h}_k \in \mathcal{R}^p$  are, respectively, the transmitted symbol and the signature spreading sequence of user  $k$ , and  $\mathbf{W}$  is the Gaussian noise with zero mean and covariance matrix  $\sigma^2 \mathbf{I}$ . Assume that the transmitted symbols of different users are independent, with  $Ex_k = 0$  and  $E|x_k|^2 = p_k$ . This model is slightly more general than that in [15], where all of the users' powers  $p_k$  are assumed to be the same.

Following [15], consider the demodulation of user 1 and use the signal-to-interference ratio (SIR) as the performance measure of linear receivers. The SIR of user 1 is defined by (see [15])

$$\beta_1 = \frac{(\mathbf{c}_1^T \mathbf{h}_1)^2 p_1}{\mathbf{c}_1^T \mathbf{c}_1 \sigma^2 + \sum_{k=2}^K (\mathbf{c}_1^T \mathbf{h}_k)^2 p_k}.$$

The minimum mean square error (MMSE) receiver minimizes the mean square error as well as maximizes the SIR for all users (see [15]). The SIR of user 1 is given by

$$\beta_1^{\text{MMSE}} = p_1 \mathbf{h}_1^T (\mathbf{H}_1 \mathbf{D}_1 \mathbf{H}_1^T + \sigma^2 \mathbf{I})^{-1} \mathbf{h}_1,$$

where

$$\mathbf{D}_1 = \text{diag}(p_2, \dots, p_n), \quad \mathbf{H}_1 = (\mathbf{h}_2, \dots, \mathbf{h}_n).$$

Assume that the  $\mathbf{h}'_k$  are i.i.d. random vectors, each consisting of i.i.d. random variables with appropriate moments. Moreover, suppose that  $p/n \rightarrow c > 0$  and  $F^{\mathbf{D}_1}(x) \rightarrow H(x)$ . Then, by Lemma 2.7 in [2] and the Helly–Bray lemma, it is not difficult to check that

$$\beta_1^{\text{MMSE}} - p_1 \int \frac{1}{x + \sigma^2} dF_{c,H}(x) \xrightarrow{i.p.} 0.$$

To judge the performance of different receivers, we may then compare the value of  $\int \frac{1}{x + \sigma^2} dF_{c,H}(x)$  with the limiting SIR of the other linear receiver. However, the awkward fact is that we usually do not have an explicit expression for  $F_{c,H}(x)$ . Thus, we may use the kernel estimate  $\int \frac{1}{x + \sigma^2} dF_n(x)$  to estimate  $\int \frac{1}{x + \sigma^2} dF_{c,H}(x)$ , by Corollary 2.

A second application: we may use  $f_n(x)$  to infer, in some way, some statistical properties of the population covariance matrix  $\mathbf{T}_n$ . Specifically speaking, by (1.3), we may evaluate the Stieltjes transform of the kernel estimator  $f_n(x)$ ,

$$(3.2) \quad m_{f_n}(z) = \int \frac{1}{x - z} f_n(x) dx, \quad z \in \mathcal{C}^+.$$

We may then obtain  $\underline{m}_{f_n}(z)$ , by (1.5). On the other hand, we conclude from (1.6) that

$$(3.3) \quad \frac{\underline{m}(z)(c - 1 - z\underline{m}(z))}{c} = \int \frac{dH(t)}{t + 1/\underline{m}(z)}.$$

Note that  $\underline{m}(z)$  has a positive imaginary part. Therefore, with notation  $z_1 = -1/\underline{m}(z)$  and  $s(z_1) = \frac{\underline{m}(z)(c-1-z\underline{m}(z))}{c}$ , we can rewrite (3.3) as

$$(3.4) \quad s(z_1) = \int \frac{dH(t)}{t - z_1}, \quad z_1 \in \mathcal{C}^+.$$

Consequently, in view of the inversion formula

$$(3.5) \quad F\{[a, b]\} = \frac{1}{\pi} \lim_{v \rightarrow 0} \int_a^b \Im m_F(u + iv) du,$$

we may recover  $H(t)$  from  $s(z_1)$  as given in (3.4). However,  $s(z_1)$  can be estimated by the resulting kernel estimate

$$(3.6) \quad \frac{\underline{m}_{f_n}(z)(c-1-z\underline{m}_{f_n}(z))}{c}.$$

Once  $H(t)$  is estimated, we may further estimate the functions of the population covariance matrix  $\mathbf{T}_n$ , such as  $\frac{1}{n} \text{tr} \mathbf{T}_n^2$ . Indeed, by the Helly–Bray lemma, we have

$$\frac{1}{n} \text{tr} \mathbf{T}_n^2 = \int t^2 dH_n(t) \xrightarrow{D} \int t^2 dH(t).$$

Thus, we may construct an estimator for  $\frac{1}{n} \text{tr} \mathbf{T}_n^2$  based on the resulting kernel estimate (3.6). We conjecture that the estimators of  $H(t)$  and the corresponding functions like  $\frac{1}{n} \text{tr} \mathbf{T}_n^2$ , obtained by the above method, are also consistent. A rigorous argument is currently being pursued.

**4. Simulation study.** In this section, we perform a simulation study to investigate the behavior of the kernel density estimators of the Marcenko and Pastur law. We consider two different populations, exponential and binomial distributions. From each population, we generate two samples with sizes  $50 \times 200$  and  $800 \times 3200$ , respectively. We can therefore form two random matrices,  $(X_{ij})_{50,200}$  and  $(X_{ij})_{800,3200}$ . The kernel is selected as

$$K(x) = (2\pi)^{-1/2} e^{-x^2/2},$$

which is the standard normal density function. The bandwidth is chosen as  $h = 0.5n^{-1/3}$  ( $n = 200, 3200$ ).

For  $(X_{ij})_{50,200}$ , the kernel density estimator is

$$\frac{1}{50 \times 200^{-2/5}} \sum_{i=1}^{50} K((x - \mu_i)/200^{-2/5}),$$

where  $\mu_i, i = 1, \dots, 50$ , are eigenvalues of  $200^{-1}(X_{ij})_{50,200}(X_{ij})_{50,200}^T$ . This curve is drawn by dot-dash lines in the first two pictures.

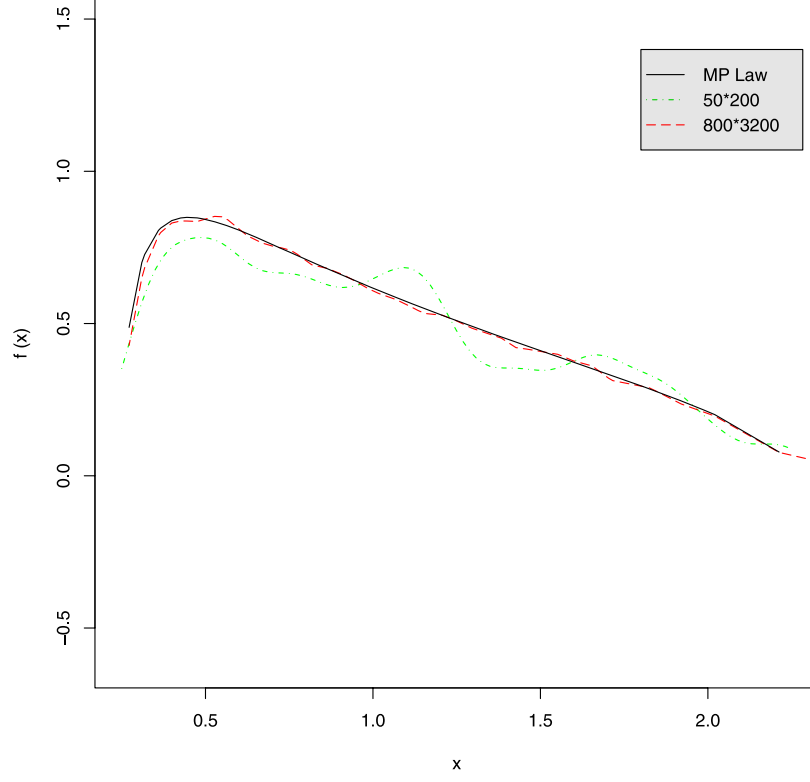


FIG. 1. Spectral density curves for sample covariance matrices  $n^{-1}(X_{ij})_{p \times n}(X_{ij})_{p \times n}^T$ ,  $X_{ij} \sim \text{exponential distribution}$ .

For  $(X_{ij})_{800,3200}$ , the kernel density estimator is

$$\frac{1}{800 \times 3200^{-2/5}} \sum_{i=1}^{800} K((x - \mu_i)/3200^{-2/5}),$$

where  $\mu_i$ ,  $i = 1, \dots, 800$ , are eigenvalues of  $3200^{-1}(X_{ij})_{800,3200}(X_{ij})_{800,3200}^T$ . This curve is drawn by dashed lines in the first two pictures.

The density function of the Marcenko and Pastur law is drawn by solid lines in the first two pictures. Here, in Figure 1, the distribution is

$$(4.1) \quad F(x) = e^{-(x+1)}, \quad x \geq -1.$$

In Figure 2, the distribution is

$$(4.2) \quad P(X = -1) = 1/2, \quad P(X = 1) = 1/2.$$

From the two figures, we see that the estimated curves fit the Marcenko and Pastur law very well. As  $n$  becomes large, the estimated curves become closer to the Marcenko and Pastur law.



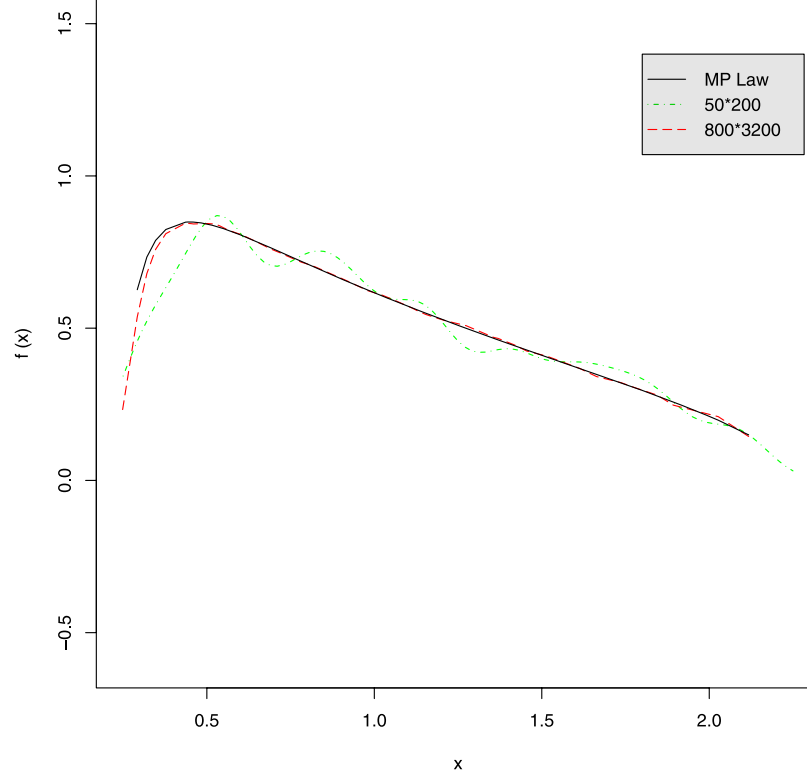


FIG. 2. Spectral density curves for sample covariance matrices  $n^{-1}(X_{ij})_{p \times n}(X_{ij})_{p \times n}^T$ ,  $X_{ij} \sim \text{binomial distribution}$ .

Finally, we consider the estimated density curves based on the following three matrices:

$$\begin{aligned} \mathbf{A}_{200} &= \frac{1}{200} \mathbf{T}_{200}^{1/2} \mathbf{X}_{50 \times 200} \mathbf{X}_{50 \times 200}^T \mathbf{T}_{200}^{1/2}, \\ \mathbf{A}_{3200} &= \frac{1}{3200} \mathbf{T}_{800}^{1/2} \mathbf{X}_{800 \times 3200} \mathbf{X}_{800 \times 3200}^T \mathbf{T}_{3200}^{1/2}, \\ \mathbf{A}_{6400} &= \frac{1}{6400} \mathbf{T}_{6400}^{1/2} \mathbf{X}_{1600 \times 6400} \mathbf{X}_{1600 \times 6400}^T \mathbf{T}_{6400}^{1/2}, \end{aligned}$$

where  $\mathbf{X}_{p \times 4p}$ ,  $p = 50, 800, 1600$ , are  $p \times 4p$  matrices whose elements are i.i.d. random variables with distribution (4.1), and  $\mathbf{T}_n = \frac{1}{4p} \mathbf{Y}_{p \times 4p} \mathbf{Y}_{p \times 4p}^T$ . Here,  $\mathbf{Y}_{p \times 4p}$  is a  $p \times 4p$  matrix consisting of i.i.d. random variables whose distributions are given by (4.2).  $\mathbf{T}_n$  and  $\mathbf{X}_{p \times 4p}$  are independent. The kernel function is the same as before. The bandwidths corresponding to the three matrices are  $0.5 \times (4p)^{-1/3}$ . In Figure 3, we present three estimated curves. The dot-dash line is based on  $\mathbf{A}_{200}$ , the dashed line on  $\mathbf{A}_{3200}$  and the solid line on  $\mathbf{A}_{6400}$ . Although, in this case, we do not know its exact formula, we can predict the limiting spectral density function from Figure 3.

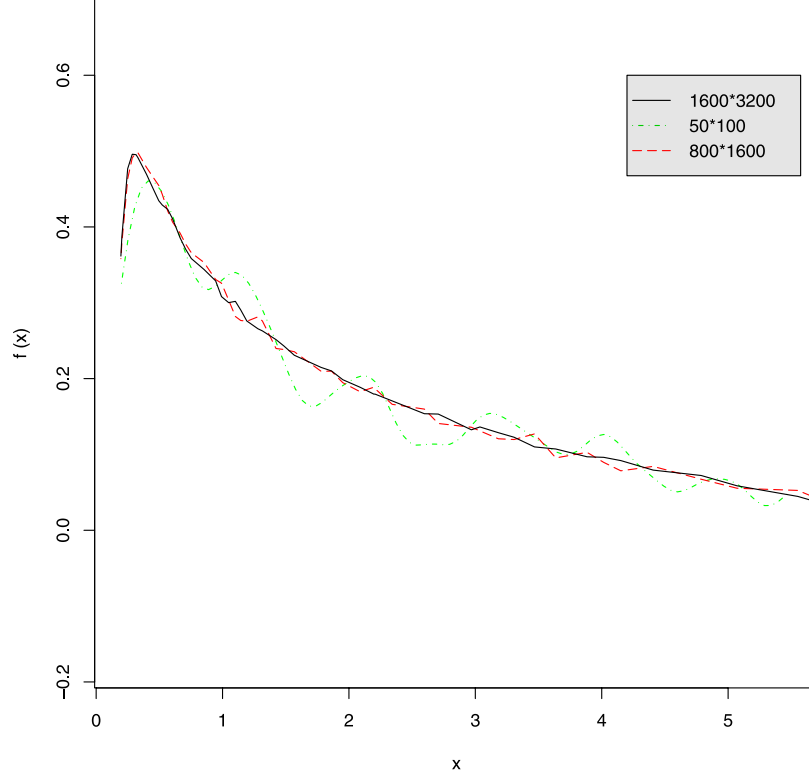


FIG. 3. Spectral density curves for sample covariance matrices  $n^{-1}\mathbf{T}_n^{1/2}(X_{ij})_{p \times n}(X_{ij})_{p \times n}^T\mathbf{T}_n^{1/2}$ ,  $X_{ij} \sim \text{exponential distribution}$ ,  $\mathbf{T}_n = n^{-1}(Y_{ij})_{p \times n}(Y_{ij})_{p \times n}^T$ ,  $Y_{ij} \sim \text{binomial distribution}$ .

In order to show that the above conclusion is reliable, we choose ten points throughout the range and calculate the mean square errors (MSEs) for the kernel density estimator at the selected ten points, based on 500 matrices,

$$\text{MSE}(x) = 500^{-1} \sum_{i=1}^{500} (f_n^{(i)}(x) - f_c(x))^2,$$

where  $f_n^{(i)}(x)$  is the kernel density estimator at  $x$  based on the  $i$ th matrix. If the limiting distribution is unknown as in the case  $\mathbf{A}_{200}$ , we use the averaged spectral density

$$\bar{f}_c(x) = 500^{-1} \sum_{i=1}^{500} f_n^{(i)}(x).$$

TABLE 1  
*MSE of spectral density curves for sample covariance matrices*  
 $n^{-1}(X_{ij})_{p \times n}(X_{ij})_{p \times n}^T$ ,  $X_{ij} \sim \text{exponential distribution}$

$x =$	<b>0.30</b>	<b>0.511</b>	<b>0.722</b>	<b>0.933</b>	<b>1.144</b>
$50 \times 200$	9.89e-2	3.21e-2	3.18e-2	3.25e-2	3.56e-2
$800 \times 3200$	3.84e-03	7.44e-5	7.28e-5	7.67e-5	7.34e-5
$x =$	<b>1.356</b>	<b>1.567</b>	<b>1.778</b>	<b>1.989</b>	<b>2.20</b>
$50 \times 200$	3.79e-2	3.18e-2	3.73e-2	2.76e-2	3.63e-2
$800 \times 3200$	7.67e-5	7.23e-5	6.88e-5	6.60e-5	6.74e-5

So, in this case,

$$\text{MSE}(x) = 500^{-1} \sum_{i=1}^{500} (f_n^{(i)}(x) - \bar{f}_c(x))^2.$$

The numerical results for the three different matrices considered in this section are presented in Tables 1, 2 and 3. The notation “e- $j$ ” in these tables means multiplication by  $10^{-j}$ . The MSEs are uniformly small. As  $n$  becomes large, the MSEs become smaller. This supports the conclusion that our proposed kernel spectral density curve is consistent.

We also conducted simulations using a wide range of bandwidths from small  $h = n^{-1/2}$  to large  $h = n^{-1/10}$ . The kernel spectral density curves seem to change rather slowly. This indicates that the kernel spectral density estimator is robust with respect to the bandwidth selection.

**5. Proofs of Theorems 1 and 2.** Throughout this section and the next, to simplify notation,  $M, M_1, \dots, M_{12}$  stand for constants which may take different values from one appearance to the next.

TABLE 2  
*MSE of spectral density curves for sample covariance matrices*  
 $n^{-1}(X_{ij})_{p \times n}(X_{ij})_{p \times n}^T$ ,  $X_{ij} \sim \text{binomial distribution}$

$x =$	<b>0.30</b>	<b>0.511</b>	<b>0.722</b>	<b>0.933</b>	<b>1.144</b>
$50 \times 200$	3.23e-1	3.14e-2	2.38e-2	2.76e-2	2.86e-2
$800 \times 3200$	5.13e-03	8.01e-5	6.05e-5	7.30e-5	6.53e-5
$x =$	<b>1.356</b>	<b>1.567</b>	<b>1.778</b>	<b>1.989</b>	<b>2.20</b>
$50 \times 200$	2.70e-2	2.44e-2	2.42e-2	2.40e-2	1.69e-2
$800 \times 3200$	6.28e-5	7.65e-5	6.14e-5	6.68e-5	1.13e-4

TABLE 3

*MSE of spectral density curves for sample covariance matrices*  
 $n^{-1}\mathbf{T}_n^{1/2}(X_{ij})_{p \times n}(X_{ij})_{p \times n}^T\mathbf{T}_n^{1/2}$ ,  $X_{ij} \sim \text{exponential distribution}$   
 $\mathbf{T}_n = n^{-1}(Y_{ij})_{p \times n}(Y_{ij})_{p \times n}^T$ ,  $Y_{ij} \sim \text{binomial distribution}$   
 $n^{-1}(X_{ij})_{p \times n}(X_{ij})_{p \times n}^T$ ,  $X_{ij} \sim \text{binomial distribution}$

$x =$	<b>0.30</b>	<b>0.511</b>	<b>0.722</b>	<b>0.933</b>	<b>1.144</b>
$50 \times 100$	1.20e-2	8.71e-3	8.58e-3	7.90e-3	8.77e-3
$800 \times 1600$	6.25e-05	4.00e-5	3.51e-5	3.19e-5	2.71e-5
$1600 \times 3200$	2.98e-5	1.83e-5	1.44e-5	1.39e-5	1.53e-5
$x =$	<b>1.356</b>	<b>1.567</b>	<b>1.778</b>	<b>1.989</b>	<b>2.20</b>
$50 \times 200$	7.91e-3	8.07e-3	8.34e-3	7.54e-3	7.17e-3
$800 \times 3200$	3.04e-5	3.10e-5	2.98e-5	2.89e-5	2.66e-5
$1600 \times 3200$	1.19e-5	1.19e-5	1.36e-5	1.29e-5	1.32e-5

5.1. *Proof of Theorem 1.* We begin by developing the following two lemmas, necessary for the argument of Theorem 1.

LEMMA 1. *Under the assumptions of Theorem 1, let  $F_{c_n, H_n}(t)$  be the distribution function obtained from  $F_{c, H}(t)$  by replacing  $c$  and  $H$  by  $c_n$  and  $H_n$ , respectively. Furthermore,  $f_{c_n, H_n}(x)$  denotes the density of  $F_{c_n, H_n}(x)$ . Then,*

$$\sup_{n, x} f_{c_n, H_n}(x) \leq M.$$

PROOF. From (3.10) in [2], we have

$$(5.1) \quad z(\underline{m}_n) = -\frac{1}{\underline{m}_n} + c_n \int \frac{t dH_n(t)}{1 + t\underline{m}_n},$$

where  $\underline{m}_n = \underline{m}_n(z) = \underline{m}_{F_{c_n, H_n}}(z)$ . Based on this expression, conclusions similar to those in Theorem 1.1 of [14] still hold if we replace  $F_{c, H}(x)$  by  $F_{c_n, H_n}(x)$  and then argue similarly with the help of [14]. For example, the equality (1.6) in Theorem 1.1 of [14] states that

$$(5.2) \quad x = -\frac{1}{\underline{m}(x)} + c \int \frac{t dH(t)}{1 + t\underline{m}(x)}.$$

Similarly, for every  $x \neq 0$  for which  $f_{c_n, H_n}(x) > 0$ ,  $\pi f_{c_n, H_n}(x)$  is the imaginary part of the unique  $\underline{m}_n(x)$  satisfying

$$(5.3) \quad x = -\frac{1}{\underline{m}_n(x)} + c_n \int \frac{t dH_n(t)}{1 + t\underline{m}_n(x)}.$$

Now, consider the imaginary part of  $\underline{m}_n(x)$ . From (5.3), we obtain

$$(5.4) \quad c_n \int \frac{t^2 dH_n(t)}{|1 + t\underline{m}_n(x)|^2} = \frac{1}{|\underline{m}_n(x)|^2}.$$

It follows from (5.3), (5.4) and Hölder's inequality that

$$\begin{aligned} |\underline{m}_n(x)| &\leq \frac{|c_n - 1|}{x} + \frac{c_n}{x} \int \frac{dH_n(t)}{|1 + t\underline{m}_n(x)|} \\ &\leq \frac{|c_n - 1|}{x} + \frac{c_n}{x} \left( \int \frac{t^2 dH_n(t)}{|1 + t\underline{m}_n(x)|^2} \int \frac{dH_n(t)}{t^2} \right)^{1/2} \\ &\leq \frac{|c_n - 1|}{x} + \frac{\sqrt{c_n}}{x|\underline{m}_n(x)|} \left( \int \frac{dH_n(t)}{t^2} \right)^{1/2}, \end{aligned}$$

where  $\int \frac{dH_n(t)}{t^2}$  is well defined because we require the support of  $F_{c,H}(x)$  to be  $[a, b]$  with  $a > 0$ . This inequality is equivalent to

$$|\underline{m}_n(x)|^2 \leq \frac{|c_n - 1|}{x} |\underline{m}_n(x)| + \frac{\sqrt{c_n}}{x} \left( \int \frac{dH_n(t)}{t^2} \right)^{1/2}.$$

It follows that

$$(5.5) \quad \sup_{n,x} |\underline{m}_n(x)| \leq M.$$

This leads to  $\sup_{n,x} f_{c_n, H_n}(x) \leq M$ .  $\square$

LEMMA 2. *Under the assumptions of Lemma 1, when  $x_n \rightarrow x$ , we have*

$$(5.6) \quad f_{c_n, H_n}(x_n) - f_{c, H}(x_n) \rightarrow 0.$$

PROOF. Obviously,  $f_{c, H}(x_n) - f_{c, H}(x) \rightarrow 0$  because  $f_{c, H}(x)$  is continuous on the interval  $[a, b]$ . Moreover, in view of (5.5), we may choose a subsequence  $n_k$  so that  $\underline{m}_{n_k}(x_{n_k})$  converges. We denote its limit by  $a(x)$ . Suppose that  $\Im(a(x)) > 0$ . Then, as in Lemma 3.3 in [14], we may argue that the limit of  $\underline{m}_n(x_n)$  exists as  $n \rightarrow \infty$ . Next, we verify that  $a(x) = \underline{m}(x)$ . By (5.3), we then have

$$x = -\frac{1}{a(x)} + c \int \frac{t dH(t)}{1 + ta(x)}$$

because, via (5.4) and Hölder's inequality,

$$\begin{aligned} &\left| \int \frac{t dH_n(t)}{1 + t\underline{m}_n(x)} - \int \frac{t dH_n(t)}{1 + ta(x)} \right| \\ &\leq |\underline{m}_n(x) - a(x)| \left( \frac{1}{c_n |\underline{m}_n(x)|^2} \int \frac{t^2 dH_n(t)}{|1 + ta(x)|^2} \right)^{1/2} \end{aligned}$$

and

$$\int \frac{t dH_n(t)}{1 + ta(x)} \rightarrow \int \frac{t dH(t)}{1 + ta(x)}.$$

Since the solution satisfying the equation (5.2) is unique,  $a(x) = \underline{m}(x)$ . Therefore,  $\underline{m}_n(x) \rightarrow \underline{m}(x)$ , which then implies that

$$(5.7) \quad f_{c_n, H_n}(x_n) - f_{c, H}(x) \rightarrow 0.$$

Now, suppose that  $\Im(a(x)) = 0$ . This implies that  $\Im(\underline{m}_n(x_n)) \rightarrow 0$  and then that  $f_{c_n, H_n}(x_n) \rightarrow 0$  because if there is another subsequence on which  $\Im(\underline{m}_n(x_n))$  converges to a positive number, then  $\underline{m}_n(x_n)$  must converge to the complex number with the positive imaginary part, by the previous argument. Next, by (1.2) and (5.1),  $\Im(\underline{m}_n(x_n + iv)) - \Im(\underline{m}(x_n + iv)) \rightarrow 0$  for any  $v > 0$ . We may then choose  $v_n \rightarrow 0$  so that  $\Im(\underline{m}_n(x_n + iv_n)) - \Im(\underline{m}(x_n + iv_n)) \rightarrow 0$  as  $n \rightarrow \infty$ . Moreover,  $\Im(\underline{m}(x_n + iv_n)) \rightarrow \Im(\underline{m}(x))$  and  $\Im(\underline{m}_n(x_n + iv_n)) - \Im(\underline{m}_n(x)) \rightarrow 0$  by Theorem 1.1 of [14] and a theorem for  $\underline{m}_n(z)$  similar to Theorem 1.1 of [14]. Therefore, in view of the continuity of  $\underline{m}_n(x)$  for  $x \neq 0$ ,  $\Im(\underline{m}(x)) = 0$  and then (5.6) holds for the case  $\Im(a(x)) = 0$ .  $\square$

We now proceed to prove Theorem 1. First, we claim that

$$(5.8) \quad \sup_x \left| f_n(x) - \frac{1}{h} \int K\left(\frac{x-t}{h}\right) dF_{c_n, H_n}(t) \right| \rightarrow 0$$

in probability. Indeed, from integration by parts and Theorem 3, we obtain

$$\begin{aligned} & E \sup_x \left| \frac{1}{h} \int K\left(\frac{x-t}{h}\right) dF^{\mathbf{A}_n}(t) - \frac{1}{h} \int K\left(\frac{x-t}{h}\right) dF_{c_n, H_n}(t) \right| \\ &= E \sup_x \left| \frac{1}{h^2} \int K'\left(\frac{x-t}{h}\right) (F^{\mathbf{A}_n}(t) - F_{c_n, H_n}(t)) dt \right| \\ &= E \sup_x \left| \frac{1}{h} \int K'(u) (F^{\mathbf{A}_n}(x - uh) - F_{c_n, H_n}(x - uh)) du \right| \\ &\leq \frac{1}{h} E \sup_x |F^{\mathbf{A}_n}(x) - F_{c_n, H_n}(x)| \int |K'(u)| du \\ &\leq \frac{M}{n^{2/5}h} \rightarrow 0. \end{aligned}$$

The next aim is to show that

$$\frac{1}{h} \int K\left(\frac{x-t}{h}\right) dF_{c_n, H_n}(t) - \frac{1}{h} \int K\left(\frac{x-t}{h}\right) dF_{c, H}(t) \rightarrow 0$$

uniformly in  $x \in [a, b]$ . This is equivalent to, for any sequence  $\{x_n, n \geq 1\}$  in  $[a, b]$  converging to  $x$ ,

$$(5.9) \quad \int K(u)(f_{c_n, H_n}(x_n - uh) - f_{c, H}(x_n - uh)) du \rightarrow 0.$$

From Theorem 1.1 of [14],  $f_{c, H}(x)$  is uniformly bounded on the interval  $[a, b]$ . Therefore, (5.9) follows from the dominated convergence theorem, Lemma 1 and Lemma 2.

Finally,

$$\begin{aligned} & \left| \frac{1}{h} \int K\left(\frac{x-t}{h}\right) dF_{c, H}(t) - f_{c, H}(x) \frac{1}{h} \int_{x-b}^{x-a} K\left(\frac{t}{h}\right) dt \right| \\ &= \left| \int_{x-b}^{x-a} (f_{c, H}(x-t) - f_{c, H}(x)) \frac{1}{h} K\left(\frac{t}{h}\right) dt \right| \\ &\leq \sup_{x \in [a, b]} \int_{|t| > \delta} \left| (f_{c, H}(x-t) - f_{c, H}(x)) \frac{1}{h} K\left(\frac{t}{h}\right) \right| dt \\ &\quad + \sup_{x \in [a, b]} \int_{|t| \leq \delta} \left| (f_{c, H}(x-t) - f_{c, H}(x)) \frac{1}{h} K\left(\frac{t}{h}\right) \right| dt \\ &\leq 2 \sup_{x \in [a, b]} f_{c, H}(x) \int_{|t| > \delta/h} |K(y)| dy \\ &\quad + \sup_{x \in [a, b]} \sup_{|t| \leq \delta} |f_{c, H}(x-t) - f_{c, H}(x)| \int \frac{1}{h} \left| K\left(\frac{t}{h}\right) \right| dt, \end{aligned}$$

which goes to zero by fixing  $\delta$  and letting  $n \rightarrow \infty$  first, and then letting  $\delta \rightarrow 0$ . On the other hand, obviously,

$$\frac{1}{h} \int_{x-b}^{x-a} K\left(\frac{t}{h}\right) dt = \int_{(x-b)/h}^{(x-a)/h} K(t) dt \rightarrow \int_{-\infty}^{+\infty} K(t) dt = 1.$$

Thus, the proof is complete.

**5.2. Proof of Theorem 2.** Denote by  $F_{c_n}(t)$  the distribution function obtained from  $F_c(t) = \int_{-\infty}^t f_c(x) dx$  with  $c$  replaced by  $c_n$ . Let  $\mathbf{S}_n = \frac{1}{n} \mathbf{X}_n \mathbf{X}_n^T$ . From integration by parts, we obtain

$$\begin{aligned} & \left| \frac{1}{h} \int K\left(\frac{x-t}{h}\right) dF^{\mathbf{S}_n}(t) - \frac{1}{h} \int K\left(\frac{x-t}{h}\right) dF_{c_n}(t) \right| \\ &= \left| \frac{1}{h^2} \int K'\left(\frac{x-t}{h}\right) (F^{\mathbf{S}_n}(t) - F_{c_n}(t)) dt \right| \\ &= \left| \frac{1}{h} \int K'(u) (F^{\mathbf{S}_n}(x-uh) - F_{c_n}(x-uh)) du \right| \end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{h} \sup_x |F^{\mathbf{S}_n}(x) - F_{c_n}(x)| \int |K'(u)| du \\
&\leq \frac{M}{\sqrt{nh}},
\end{aligned}$$

where the last step uses Theorem 1.2 in [6]. We next prove that

$$\sup_x \left| \frac{1}{h} \int K\left(\frac{x-t}{h}\right) dF_{c_n}(t) - \frac{1}{h} \int K\left(\frac{x-t}{h}\right) dF_c(t) \right| \rightarrow 0.$$

It suffices to prove that

$$(5.10) \quad \sup_x |f_{c_n}(x) - f_c(x)| \rightarrow 0,$$

where  $f_{c_n}(x)$  stands for the density of  $F_{c_n}(x)$ .

Note that when  $c < 1$ ,

$$f_{c_n}(x) - f_c(x) = \frac{\sqrt{(x-a(c_n))(b(c_n)-x)}}{2\pi c_n x} - \frac{\sqrt{(x-a(c))(b(c)-x)}}{2\pi c x},$$

where

$$a(c) = (1 - \sqrt{c})^2, \quad b(c) = (1 + \sqrt{c})^2,$$

and  $a(c_n)$  and  $b(c_n)$  are obtained from  $a(c)$  and  $b(c)$  by replacing  $c$  with  $c_n$ , respectively. It is then a simple matter to verify that (5.10) holds for  $x \in [a(c), b(c)]$ .

Finally, as in Theorem 1, one may prove that

$$\sup_x \left| \frac{1}{h} \int K\left(\frac{x-t}{h}\right) dF_c(t) - f_c(x) \right| \rightarrow 0.$$

Thus, the proof is complete.

5.3. *Proof of Corollary 1.* The result follows from Theorem 1 in [11].

## 6. Proof of Theorem 3.

6.1. *Summary of argument.* The strategy is to use Corollary 2.2 and Lemma 7.1 in [6]. To this end, a key step is to establish an upper bound for  $|b_1|$ , defined below. Note that in a suitable interval for  $z$  with a well-chosen imaginary part  $v$ , the absolute value of the expectation of the Stieltjes transform of  $F_{\mathbf{A}_n}$ ,  $|Em_n(z)|$ , is bounded. Moreover, for such  $v$ , when  $n \rightarrow \infty$ , the difference between  $b_1$  and its alternative expression involving  $Em_n(z)$ ,  $\rho_n$  [given in (6.13)], converges to zero with some convergence rate. Therefore, we may argue that  $|b_1|$  is bounded. Once this is done, we further develop a convergence rate of  $m_n(z) - Em_n(z)$  using a martingale decomposition, and a convergence rate of the difference between  $Em_n(z)$  and its corresponding limit using a recurrence approach.

We begin by giving some notation. Define  $\mathbf{A}(z) = \mathbf{A}_n - z\mathbf{I}$ ,  $\mathbf{A}_j(z) = \mathbf{A}(z) - \mathbf{s}_j \mathbf{s}_j^T$  and  $\mathbf{s}_j = \mathbf{T}_n^{1/2} \mathbf{x}_j$ , with  $\mathbf{x}_j$  being the  $j$ th column of  $\mathbf{X}_n$ . Let



$E_j = E(\cdot | \mathbf{s}_1, \dots, \mathbf{s}_j)$  and let  $E_0$  denote the expectation. Moreover, introduce

$$\begin{aligned}\beta_j &= \frac{1}{1 + \mathbf{s}_j^T \mathbf{A}_j^{-1}(z) \mathbf{s}_j}, & \hat{\beta}_j &= \frac{1}{1 + n^{-1} \text{tr} \mathbf{T}_n \mathbf{A}_j^{-1}(z)}, \\ \eta_j &= \mathbf{s}_j^T \mathbf{A}_j^{-1}(z) \mathbf{s}_j - \frac{1}{n} \text{tr} \mathbf{A}_j^{-1}(z) \mathbf{T}_n, & b_1 &= \frac{1}{1 + n^{-1} E \text{tr} \mathbf{T}_n \mathbf{A}_1^{-1}(z)}, \\ m_n(z) &= \int \frac{dF_{\mathbf{A}_n}(x)}{x - z}, & m_n^0(z) &= \int \frac{dF_{c_n, H_n}(x)}{x - z}, \\ \underline{m}_n(z) &= \int \frac{dF_{\mathbf{B}_n}(x)}{x - z}, & \underline{m}_n^0(z) &= \int \frac{dF_{c_n, H_n}(x)}{x - z}\end{aligned}$$

and

$$\xi_1 = \mathbf{s}_1^T \mathbf{A}_1^{-1}(z) \mathbf{s}_1 - \frac{1}{n} E \text{tr} \mathbf{A}_1^{-1}(z) \mathbf{T}_n.$$

Here,  $F_{c_n, H_n}(x)$  is obtained from  $F_{c, H}(x)$  by replacing  $c$  and  $H$  by  $c_n$  and  $H_n$ , respectively.

Let  $\Delta_n = \sup_x |EF^{\mathbf{A}_n}(x) - F_{c_n, H_n}(x)|$  and  $v_0 = \max\{\gamma \Delta_n, M_1 n^{-2/5}\}$  with  $0 < \gamma < 1$  to be chosen later and  $M_1$  an appropriate constant. As in Lemma 3.1 and Lemma 3.2 in [14], we obtain, for  $u \in [a, b]$  and  $v_0 \leq v \leq 1$ ,

$$(6.1) \quad |\underline{m}_n^0(z)| \leq M, \quad |m_n^0(z)| \leq M,$$

where the bound for  $|m_n^0(z)|$  is obtained with the help of (1.5). Using integration by parts, we have, for  $v > v_0$ ,

$$\begin{aligned}|Em_n(z) - m_n^0(z)| &= \left| \int_{-\infty}^{+\infty} \frac{1}{x - z} d(EF^{\mathbf{A}_n}(x) - F_{c_n, H_n}(x)) \right| \\ &= \left| \int_{-\infty}^{+\infty} \frac{EF^{\mathbf{A}_n}(x) - F_{c_n, H_n}(x)}{(x - z)^2} dx \right| \leq \frac{\pi \Delta_n}{v} \leq \frac{\pi}{\gamma}.\end{aligned}$$

This implies that

$$(6.2) \quad |Em_n(z)| \leq M, \quad |E\underline{m}_n(z)| \leq M,$$

where the bound for  $|E\underline{m}_n(z)|$  is obtained from an equality similar to (1.5), noting that  $\Re z \geq a$ . It is readily observed that  $|\hat{\beta}_j|$  and  $|\beta_j|$  are both bounded by  $|z|/v$  (see (3.4) in [2]) and that Lemma 2.10 in [2] yields

$$(6.3) \quad |\beta_j \mathbf{s}_j^T \mathbf{A}_j^{-2}(z) \mathbf{s}_j| \leq v^{-1},$$

which gives

$$(6.4) \quad |\text{tr}(\mathbf{A} - z\mathbf{I})^{-1} - \text{tr}(\mathbf{A}_k - z\mathbf{I})^{-1}| \leq v^{-1}.$$

This, together with (6.2), gives, for  $v > v_0$ ,

$$(6.5) \quad \left| \frac{1}{n} E \text{tr} \mathbf{A}_1^{-1}(z) \right| \leq M.$$

In the subsequent subsections, we will assume that  $z = u + iv$  with  $v \geq v_0$  and  $u \in [a, b]$ .

6.2. *Bounds for  $n^{-2}E|\operatorname{tr} \mathbf{A}^{-1}(z) - E \operatorname{tr} \mathbf{A}^{-1}(z)|^2$  and  $E|\beta_1|^2$ .*

LEMMA 3. *If  $|b_1| \leq M$ , then, for  $v > M_1 n^{-2/5}$ ,*

$$(6.6) \quad \frac{1}{n^2}E|\operatorname{tr} \mathbf{A}^{-1}(z) - E \operatorname{tr} \mathbf{A}^{-1}(z)|^2 \leq \frac{M}{n^2 v^3}.$$

PROOF.

$$\begin{aligned} & \frac{1}{n} \operatorname{tr} \mathbf{A}^{-1}(z) - E \operatorname{tr} \mathbf{A}^{-1}(z) \\ &= \frac{1}{n} \sum_{j=1}^n (E_j \operatorname{tr} \mathbf{A}^{-1}(z) - E_{j-1} \operatorname{tr} \mathbf{A}^{-1}(z)) \\ &= \frac{1}{n} \sum_{j=1}^n E_j (\operatorname{tr} \mathbf{A}^{-1}(z) - \operatorname{tr} \mathbf{A}_j^{-1}(z)) - E_{j-1} \operatorname{tr} (\operatorname{tr} \mathbf{A}^{-1}(z) - \operatorname{tr} \mathbf{A}_j^{-1}(z)) \\ &= \frac{1}{n} \sum_{j=1}^n (E_j - E_{j-1}) (\beta_j \mathbf{s}_j^T \mathbf{A}_j^{-2}(z) \mathbf{s}_j) \\ &= \frac{1}{n} \sum_{j=1}^n (E_j - E_{j-1}) \left[ b_1 \left( \mathbf{s}_j^T \mathbf{A}_j^{-2}(z) \mathbf{s}_j - \frac{1}{n} \operatorname{tr} \mathbf{A}_j^{-2}(z) \mathbf{T}_n \right) \right. \\ & \quad \left. + b_1 \beta_j \mathbf{s}_j^T \mathbf{A}_j^{-2}(z) \mathbf{s}_j \xi_j \right], \end{aligned}$$

where the last step uses the fact that

$$(6.7) \quad \beta_j = b_1 - b_1 \beta_j \xi_j.$$

Lemma 2.7 in [2] then gives

$$\begin{aligned} & E \left| \frac{1}{n} \sum_{j=1}^n (E_j - E_{j-1}) \left( \mathbf{s}_j^T \mathbf{A}_j^{-2}(z) \mathbf{s}_j - \frac{1}{n} \operatorname{tr} \mathbf{A}_j^{-2}(z) \mathbf{T}_n \right) \right|^2 \\ & \leq \frac{M}{n^2} \sum_{j=1}^n E \left| \left( \mathbf{s}_j^T \mathbf{A}_j^{-2}(z) \mathbf{s}_j - \frac{1}{n} \operatorname{tr} \mathbf{A}_j^{-2}(z) \mathbf{T}_n \right) \right|^2 \\ & \leq \frac{M}{n^2} \sum_{j=1}^n E \frac{1}{n^2} \operatorname{tr} \mathbf{A}_1^{-2}(z) \mathbf{T}_n \mathbf{A}_1^{-2}(\bar{z}) \mathbf{T}_n \\ & \leq \frac{\lambda_{\max}^2(\mathbf{T}_n)}{n^3 v^2} E \operatorname{tr} \mathbf{A}_1^{-1}(z) \mathbf{A}_1^{-1}(\bar{z}) \leq \frac{M}{n^2 v^3} \end{aligned}$$

because, via (6.5),

$$(6.8) \quad \frac{1}{n} E \operatorname{tr} \mathbf{A}_1^{-1}(z) \mathbf{A}_1^{-1}(\bar{z}) = \frac{1}{v} \Im \left( \frac{1}{n} E \operatorname{tr} \mathbf{A}_1^{-1}(z) \right) \leq \frac{M}{v}.$$

Using (6.3) and Lemma 2.7 in [2], we similarly have

$$\begin{aligned} E \left| \frac{1}{n} \sum_{j=1}^n (E_j - E_{j-1}) \beta_j \mathbf{s}_j^T \mathbf{A}_j^{-2}(z) \mathbf{s}_j \xi_j \right|^2 \\ \leq \frac{M}{n^2 v^3} + \frac{M}{n^3 v^2} E |\operatorname{tr} \mathbf{A}^{-1}(z) - E \operatorname{tr} \mathbf{A}^{-1}(z)|^2. \end{aligned}$$

Summarizing the above, we have proven that

$$\left( 1 - \frac{M}{nv^2} \right) \frac{1}{n^2} E |\operatorname{tr} \mathbf{A}^{-1}(z) - E \operatorname{tr} \mathbf{A}^{-1}(z)|^2 \leq \frac{M}{n^2 v^3},$$

which implies Lemma 3 by choosing an appropriate  $M_1$  such that  $\frac{M}{nv^2} < \frac{1}{2}$ .  $\square$

LEMMA 4. *If  $|b_1| \leq M$ , then, for  $v > M_1 n^{-2/5}$ ,*

$$(6.9) \quad \frac{1}{n^4} E |\operatorname{tr} \mathbf{A}^{-1}(z) - E \operatorname{tr} \mathbf{A}^{-1}(z)|^4 \leq \frac{M}{n^4 v^6}.$$

PROOF. Lemma 4 is obtained by repeating the argument of Lemma 3 and applying

$$\begin{aligned} E \left( \frac{1}{n} \operatorname{tr} \mathbf{A}_1^{-2}(z) \mathbf{T}_n \mathbf{A}_1^{-2}(\bar{z}) \mathbf{T}_n \right)^2 \\ \leq \frac{\lambda_{\max}^4(\mathbf{T}_n)}{n^2 v^6} E |\operatorname{tr} \mathbf{A}_1^{-1}(z) - E \operatorname{tr} \mathbf{A}_1^{-1}(z)|^2 + \frac{\lambda_{\max}^4(\mathbf{T}_n)}{n^2 v^6} |E \operatorname{tr} \mathbf{A}_1^{-1}(z)|^2 \\ \leq \frac{M}{v^6}. \quad \square \end{aligned}$$

LEMMA 5. *If  $|b_1| \leq M$ , then there is some constant  $M_2$  such that for  $v \geq M_2 n^{-2/5}$ ,*

$$E |\beta_1|^2 \leq M.$$

PROOF. By (6.7), we have

$$\beta_j = b_1 - b_1^2 \xi_j + b_1^2 \beta_j \xi_j$$

and

$$(6.10) \quad \begin{aligned} E |\xi_1(z)|^4 &\leq M E |\eta_1(z)|^4 + M n^{-4} E |\operatorname{tr} \mathbf{A}_1^{-1}(z) \mathbf{T}_n - E \operatorname{tr} \mathbf{A}_1^{-1}(z) \mathbf{T}_n|^4 \\ &\leq \frac{M}{n^2 v^2} + \frac{M}{n^4 v^6} \end{aligned}$$

because repeating the argument of Lemma 3 and Lemma 4 yields

$$(6.11) \quad E \left| \frac{1}{n} \operatorname{tr} \mathbf{D} \mathbf{A}_1^{-1}(z) - E \frac{1}{n} \operatorname{tr} \mathbf{D} \mathbf{A}_1^{-1}(z) \right|^4 \leq \frac{M}{n^4 v^6 \|\mathbf{D}\|^4}$$

for a fixed matrix  $\mathbf{D}$ . It follows that

$$E|\beta_1|^2 \leq |b_1|^2 + |b_1|^4 E|\xi_1|^2 + \frac{|b_1|^4}{v} (E|\beta_1|^2 E|\xi_1|^4)^{1/2},$$

which gives

$$E|\beta_1|^2 \leq M + \frac{M}{nv} + \frac{M}{nv^2} (E|\beta_1|^2)^{1/2}.$$

Solving this inequality gives Lemma 5.  $\square$

6.3. *A bound for  $b_1(z)$ .* By (6.7) and

$$(6.12) \quad 1 - c_n - z c_n m_n(z) = \frac{1}{n} \sum_{j=1}^n \beta_j$$

(see the equality above (2.2) in [13]), we get

$$(6.13) \quad b_1 = 1 - c_n - z c_n E m_n(z) + \rho_n,$$

where

$$\rho_n = b_1 E(\beta_1 \xi_1).$$

LEMMA 6. *If  $|b_1| \leq M$ , then there is some constant  $M_3$  such that for  $v \geq M_3 n^{-2/5}$ ,*

$$|\rho_n| \leq \frac{M}{nv}.$$

PROOF. Lemma 5 and (6.10) ensure that

$$|E[\beta_1(z) \xi_1(z)]| = |b_1(z) E[\beta_1(z) \xi_1^2]| \leq M (E|\beta_1(z)|^2 E|\xi_1|^4)^{1/2} \leq \frac{M}{nv}.$$

Thus, Lemma 6 is proved.  $\square$

LEMMA 7. *If  $\Im(z + \rho_n) \geq 0$ , then there exists a positive constant  $c$  depending on  $\gamma, a, b$  such that*

$$|b_1| \leq M.$$

PROOF. Consider the case  $\Im(E m_n(z)) \geq v > 0$  first. It follows from (6.13) and the assumption that

$$\begin{aligned} \Im(c_n + z + z c_n E m_n(z) - 1) &\geq -\Im(b_1) \\ &= -|b_1|^2 \Im(1 + n^{-1} E \operatorname{tr} \mathbf{A}^{-1}(\bar{z})). \end{aligned}$$

Note that

$$\begin{aligned}
 & \Im(c_n + z + z c_n E m_n(z) - 1) \\
 (6.14) \quad &= v + v c_n \int \frac{x}{|x - z|^2} dF_{n2}(x) \\
 &= v + c_n [v \Re(E m_n(z)) + u \Im(E m_n(z))] > 0.
 \end{aligned}$$

Thus, we have

$$\begin{aligned}
 |b_1|^2 &\leq \frac{v + c_n [v \Re(E m_n(z)) + u \Im(E m_n(z))]}{c_n \Im(E m_n(z))} \\
 (6.15) \quad &\leq \frac{[1 + c_n |\Re(E m_n(z))| + c_n u] \Im(E m_n(z))}{c_n \Im(E m_n(z))} \\
 &\leq 1/c_n + M + b.
 \end{aligned}$$

Next, consider the case  $\Im(E m_n(z)) < v$ . Note that for  $u \in [a, b]$ ,

$$(6.16) \quad |\Im(E m_n(z))| \geq \frac{v}{M + v^2}.$$

This, together with (6.15), gives

$$|b_1|^2 \leq \frac{(M + v^2)[1 + c_n(|\Re(E m_n(z))| + u)]v}{c_n M v} \leq \frac{1 + c_n[|\Re(E m_n(z))| + u]}{c_n M}. \quad \square$$

LEMMA 8. *There is some constant  $M_4$  such that, for any  $v \geq M_4 n^{-2/5}$ ,*

$$\Im(z + \rho_n) > 0.$$

PROOF. First, we claim that

$$(6.17) \quad \Im(z + \rho_n) \neq 0.$$

If not,  $\Im(z + \rho_n) = 0$  implies that

$$(6.18) \quad |\rho_n| \geq |\Im(\rho_n)| = v.$$

On the other hand, if  $\Im(z + \rho_n) = 0$ , then we then conclude from Lemma 7 and Lemma 6 that

$$|\rho_n| \leq \frac{M}{nv}.$$

Thus, recalling that  $v \geq M_4 n^{-2/5}$ , we may choose an appropriate constant  $M_4$  so that

$$|\rho_n| \leq \frac{v}{3},$$

which contradicts (6.18). Therefore, (6.17) holds.

Next, note that

$$\Im(z + zn^{-1}E \operatorname{tr} \mathbf{A}_1^{-1}(z)) \geq v, \quad \Im(z + zn^{-1}E \operatorname{tr} \mathbf{A}^{-1}(z)) \geq v.$$

Therefore, when taking  $v = 1$ ,

$$|b_1(z)| \leq \frac{|z|}{v} \leq M, \quad |b(z)| \leq \frac{|z|}{v} \leq M.$$

It follows from Lemma 7 and Lemma 6 that

$$|\rho_n| \leq \frac{M}{n},$$

which implies that for  $n$  large and  $v = 1$ ,

$$(6.19) \quad \Im(z + \rho_n) > 0.$$

This, together with (6.17) and continuity of the function, ensures that (6.19) holds for  $1 \geq v \geq M_3 n^{-2/5}$ . Thus, the proof of Lemma 8 is complete.  $\square$

**6.4. Convergence of expected value.** Based on Lemma 7 and Lemma 8,  $|b_1| \leq M$  and therefore all results in Section 6.2 remain true for  $v \geq M n^{-2/5}$  with some appropriate positive constant  $M$ .

Set  $\mathbf{F}^{-1}(z) = (E\mathbf{m}_n \mathbf{T}_n + \mathbf{I})^{-1}$  and then write (see (5.2) in [2])

$$(6.20) \quad c_n \int \frac{dH_n(t)}{1 + tE\mathbf{m}_n} + z c_n E(m_n(z)) = D_n,$$

where

$$D_n = E\beta_1 \left[ \mathbf{s}_1^T \mathbf{A}_1^{-1}(z) \mathbf{F}^{-1}(z) \mathbf{s}_1 - \frac{1}{n} E(\operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{T}_n \mathbf{A}^{-1}(z)) \right].$$

It follows that (see (3.20) in [2])

$$(6.21) \quad \begin{aligned} & E\mathbf{m}_n(z) - \mathbf{m}_n^0(z) \\ &= \mathbf{m}_n^0(z) E\mathbf{m}_n \omega_n / \left( 1 - c_n E\mathbf{m}_n \mathbf{m}_n^0 \int \frac{t^2 dH_n(t)}{(1 + tE\mathbf{m}_n)(1 + t\mathbf{m}_n^0)} \right), \end{aligned}$$

where  $\omega_n = -D_n/E\mathbf{m}_n$ .

Applying (6.7), we obtain

$$\begin{aligned} D_n &= b_1 E \left[ \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{T}_n \mathbf{A}_1^{-1}(z) - \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{T}_n \mathbf{A}^{-1}(z) \right] \\ &\quad - E \left[ b_1 \beta_1 \xi_1 \left( \mathbf{s}_1^T \mathbf{A}_1^{-1}(z) \mathbf{F}^{-1}(z) \mathbf{s}_1 - \frac{1}{n} E(\operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{T}_n \mathbf{A}^{-1}(z)) \right) \right]. \end{aligned}$$

We now investigate  $D_n$ . We conclude from (6.3) and Hölder's inequality that

$$(6.22) \quad \begin{aligned} & \left| \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{T}_n \mathbf{A}_1^{-1}(z) - \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{T}_n \mathbf{A}^{-1}(z) \right| \\ & \leq \frac{M}{nv^{3/2}} \left( \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{F}^{-1}(\bar{z}) \right)^{1/2}. \end{aligned}$$

Let  $\zeta_1 = \mathbf{s}_1^T \mathbf{A}_1^{-1}(z) \mathbf{F}^{-1}(z) \mathbf{s}_1 - \frac{1}{n} (\operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{T}_n \mathbf{A}_1^{-1}(z))$ . By (6.8) and Hölder's inequality, we have

$$|Eb_1^2 \xi_1 \zeta_1| = |b_1^2 E \eta_1 \zeta_1| \leq \frac{M}{nv^{3/2}} \left( \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{F}^{-1}(\bar{z}) \right)^{1/2}$$

and by Lemma 5, Lemma 4, (6.10), (6.23) and Hölder's inequality, we have

$$\begin{aligned} & E|b_1^2 \beta_1 \xi_1^2 \zeta_1| \\ & \leq M(E|\beta_1|^2)^{1/2} (E|\eta_1|^8 E|\zeta_1|^4)^{1/4} \\ & \quad + M(E|\beta_1|^2)^{1/2} \\ & \quad \times \left( E \left[ \left| \frac{1}{n} \operatorname{tr} \mathbf{A}_1^{-1}(z) \mathbf{T}_n - E \frac{1}{n} \operatorname{tr} \mathbf{A}_1^{-1}(z) \mathbf{T}_n \right|^4 E(|\zeta_1|^2 |\mathbf{A}_1^{-1}(z)|) \right] \right)^{1/2} \\ & \leq \frac{M}{nv^{3/2}}, \end{aligned}$$

where we also use (6.11) and the fact that, via Lemma 2.11 in [2],

$$(6.23) \quad \|\mathbf{F}^{-1}(z)\| \leq \frac{M}{v}.$$

These, together with (6.7), give

$$(6.24) \quad \begin{aligned} & |Eb_1 \beta_1 \xi_1 \zeta_1| \leq |Eb_1^2 \xi_1 \zeta_1| + |Eb_1^2 \beta_1 \xi_1^2 \zeta_1| \\ & \leq \frac{M}{nv^{3/2}} + \frac{M}{nv^{3/2}} \left( \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{F}^{-1}(\bar{z}) \right)^{1/2}. \end{aligned}$$

Similarly, by (6.11), we may get

$$(6.25) \quad \left| Eb_1 \beta_1 \xi_1 \left( \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{T}_n \mathbf{A}_1^{-1}(z) - E \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{T}_n \mathbf{A}_1^{-1}(z) \right) \right| \leq \frac{M}{nv^{3/2}}.$$

In view of (6.22), we have

$$\begin{aligned} & E \left| b_1 \beta_1 \xi_1 \left( E \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{T}_n \mathbf{A}_1^{-1}(z) - E \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{T}_n \mathbf{A}^{-1}(z) \right) \right| \\ & \leq \frac{M}{nv} \left( \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{F}^{-1}(\bar{z}) \right)^{1/2}. \end{aligned}$$

Summarizing the above gives

$$(6.26) \quad |D_n| \leq \frac{M}{nv^{3/2}} + \frac{M}{nv^{3/2}} \left( \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{F}^{-1}(\bar{z}) \right)^{1/2}.$$

Now, considering the imaginary part of (6.20), we may conclude that

$$(6.27) \quad c_n \int \frac{t}{|1 + tE\underline{m}_n|^2} dH_n(t) \leq \frac{|\Im(zc_n E(m_n(z)))|}{\Im(E\underline{m}_n)} + \frac{|D_n|}{\Im(E\underline{m}_n)}.$$

Formulas (6.16), (6.2) and an equality similar to (1.5) ensure that

$$(6.28) \quad \frac{|\Im(zc_n E(m_n(z)))|}{\Im(E\underline{m}_n)} \leq \frac{u\Im(E\underline{m}_n) + v|\Re(E\underline{m}_n)|}{\Im(E\underline{m}_n)} \leq M$$

and that

$$\frac{|D_n|}{\Im(E\underline{m}_n)} \leq \frac{M}{nv^{5/2}} + \frac{M}{nv^{5/2}} \left( \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{F}^{-1}(\bar{z}) \right)^{1/2}.$$

It follows that

$$(6.29) \quad \begin{aligned} c_n \int \frac{t}{|1 + tE\underline{m}_n|^2} dH_n(t) \\ \leq M + \frac{M}{nv^{5/2}} + \frac{M}{nv^{5/2}} \left( \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{F}^{-1}(\bar{z}) \right)^{1/2}, \end{aligned}$$

which implies that

$$\begin{aligned} \left| \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{F}^{-1}(\bar{z}) \right| &= \int \frac{dH_n(t)}{|1 + tE\underline{m}_n|^2} \leq \frac{1}{\lambda_{\min}(\mathbf{T}_n)} \int \frac{t dH_n(t)}{|1 + tE\underline{m}_n|^2} \\ &\leq M + \frac{M}{nv^{5/2}} + \frac{M}{nv^{5/2}} \left( \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{F}^{-1}(\bar{z}) \right)^{1/2}. \end{aligned}$$

This inequality yields

$$(6.30) \quad \left| \frac{1}{n} \operatorname{tr} \mathbf{F}^{-1}(z) \mathbf{F}^{-1}(\bar{z}) \right| \leq M.$$

This, together with (6.26), ensures that

$$(6.31) \quad |D_n| \leq \frac{M}{nv^{3/2}}.$$

Next, we prove that

$$(6.32) \quad \inf_{n,z} |E\underline{m}_n(z)| > M > 0.$$

To this end, by (6.13) and an equality similar to (1.5), we have

$$(6.33) \quad b_1 = -zE\underline{m}_n(z) + \rho_n.$$



In view of Lemma 6 and (6.33), to prove (6.32), it is thus sufficient to show that

$$(6.34) \quad \left| \frac{1}{n} E \operatorname{tr} \mathbf{A}_1^{-1}(z) \mathbf{T}_n \right| \leq M.$$

Suppose that (6.34) is not true. There then exist subsequences  $n_k$  and  $z_k \rightarrow z_0 \neq 0$  such that  $|\frac{1}{n} E \operatorname{tr} \mathbf{A}_1^{-1}(z) \mathbf{T}_n| \rightarrow \infty$  on the subsequences  $n_k$  and  $z_k$ , which, together with (6.33) and Lemma 6, implies that  $E \underline{m}_n(z) \rightarrow 0$  on such subsequences. This, together with (6.30), ensures that on such subsequences

$$c_n \int \frac{dH_n(t)}{1 + t E \underline{m}_n} \rightarrow c,$$

which, via an equality similar to (1.5), further implies that on such subsequences,

$$(6.35) \quad c_n \int \frac{dH_n(t)}{1 + t E \underline{m}_n} + z c_n E(m_n(z)) \rightarrow 1.$$

But, on the other hand, by (6.31) and (6.20),

$$c_n \int \frac{dH_n(t)}{1 + t E \underline{m}_n} + z c_n E(m_n(z)) \rightarrow 0,$$

which contradicts (6.35). Therefore, (6.34) and, consequently, (6.32) hold.

It follows from (6.32) and (6.31) that for  $v > M_8 n^{-2/5}$ ,

$$(6.36) \quad |\omega_n| \leq \frac{M}{n v^{3/2}} \leq v,$$

where we may choose an appropriate  $M_8$ . Moreover, since (1.6) holds when  $\underline{m}$  is replaced by  $\underline{m}_n^0$ , considering the imaginary parts of both sides of the equality, we obtain

$$v = \frac{\Im(\underline{m}_n^0)}{|\underline{m}_n^0|^2} - c_n \Im(\underline{m}_n^0) \int \frac{t^2 dH_n(t)}{|1 + t \underline{m}_n^0|^2},$$

which implies that

$$c_n \Im(\underline{m}_n^0) \int \frac{t^2 dH_n(t)}{|1 + t \underline{m}_n^0|^2} \leq M.$$

It follows that

$$\left( \left( c_n \Im(\underline{m}_n^0) \int \frac{t^2 dH_n(t)}{|1 + t \underline{m}_n^0|^2} \right) / \left( v + c_n \Im(\underline{m}_n^0) \int \frac{t^2 dH_n(t)}{|1 + t \underline{m}_n^0|^2} \right) \right)^{1/2} \leq 1 - Mv.$$

Applying this and (6.36), as in (3.21) in [2], we may conclude that

$$(6.37) \quad \left| 1 - c_n E \underline{m}_n \underline{m}_n^0 \int \frac{t^2 dH_n(t)}{(1 + t E \underline{m}_n)(1 + t \underline{m}_n^0)} \right| \geq Mv.$$

This, together with (6.21) and (6.31), yields

$$(6.38) \quad |E\bar{m}_n(z) - \bar{m}_n^0(z)| \leq \frac{M}{nv^{5/2}}.$$

6.5. *Convergence rate of  $EF^{\mathbf{A}_n}$  and  $F^{\mathbf{A}_n}$ .* As in Theorem 1.1 in [14],  $f_{c_n, H_n}$  is continuous. Therefore,

$$\frac{1}{v\pi} \sup_{x \in [a+1/2\varepsilon, b-1/2\varepsilon]} \int_{|y| < 2vM} |F_{c_n, H_n}(x+y) - F_{c_n, H_n}(x)| dy \leq Mv,$$

where  $\varepsilon > vM_{11}$ . Lemma 2.1 in [5] or Lemma 2.1 and Corollary 2.2 in [6] are then applicable in our case.

First, consider  $EF^{\mathbf{A}_n}$ . For  $v \geq v_0$ , by Corollary 2.2 in [6], (6.38), we obtain, after integration in  $u$  and  $v$ ,

$$(6.39) \quad \Delta_n \leq \frac{M}{n} + M_9 v_0 + \frac{M_{10}}{nv_0^{3/2}},$$

where we set  $V$ , given in Corollary 2.2 in [6], equal to one and also use the fact that  $|E\bar{m}_n(z') - \bar{m}_n^0(z')| = O(n^{-1})$  with  $z' = u + iV$  (see Section 4 in [3]). If  $v_0 = M_1 n^{-2/5}$ , then (6.39) gives  $|\Delta_n| \leq M/n^{2/5}$ . If  $v_0 = \gamma \Delta_n$ , then we choose  $\gamma = (2M_9)^{-1}$  (here one should note that  $M_{10}$  depends on  $\gamma$ , but  $M_9$  does not depend on  $\gamma$ ). Again, (6.39) gives

$$(6.40) \quad |\Delta_n| \leq M/n^{2/5}.$$

This completes the proof of (2.10).

Now, consider the convergence rate of  $F^{\mathbf{A}_n}$ . It follows from Cauchy's inequality that

$$n^{-1} |\operatorname{tr} \mathbf{A}^{-2}(z) - E \operatorname{tr} \mathbf{A}^{-2}(z)| \leq \frac{M}{v} \sup_{z_1 \in \mathcal{C}_v} n^{-1} |\operatorname{tr} \mathbf{A}^{-1}(z_1) - E \operatorname{tr} \mathbf{A}^{-1}(z_1)|,$$

where  $\mathcal{C}_v = \{z_1 : |z - z_1| = v_0/3\}$ . This, together with Lemma 3, ensures that

$$(6.41) \quad En^{-1} |\operatorname{tr} \mathbf{A}^{-2}(z) - E \operatorname{tr} \mathbf{A}^{-2}(z)| \leq \frac{M}{nv^{5/2}}.$$

Equation (2.11) then follows from (6.41), Lemma 3, the argument leading to (6.40) and Lemma 7.1 in [6].

**Acknowledgments.** The authors would like to thank the Editor, an Associate Editor and a referee for their constructive comments which helped to improve this paper considerably.

## REFERENCES

- [1] BAI, Z. D. (1993). Convergence rate of expected spectral distributions of large random matrices. Part II. Sample covariance matrices. *Ann. Probab.* **21** 649–672. [MR1217560](#)

- [2] BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large dimensional random matrices. *Ann. Probab.* **26** 316–345. [MR1617051](#)
- [3] BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. [MR2040792](#)
- [4] BAI, Z. D. and SILVERSTEIN, J. W. (2006). *Spectral Analysis of Large Dimensional Random Matrices. Mathematics Monograph Series 2*. Science Press, Beijing.
- [5] GÖTZE, F. and TIKHOMIROV, A. (2003). Rate of convergence to the semi-circular law. *Probab. Theory Related. Fields* **127** 228–276. [MR2013983](#)
- [6] GÖTZE, F. and TIKHOMIROV, A. (2004). Rate of convergence in probability to the Marčenko–Pastur law. *Bernoulli* **10** 503–548. [MR2061442](#)
- [7] HALL, P. (1984). An optimal property of kernel estimators of a probability density. *J. Roy. Statist. Soc. Ser. B* **1** 134–138. [MR0745225](#)
- [8] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution for some sets of random matrices. *Math. USSR-Sb.* **1** 457–483.
- [9] PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.* **33** 1065–1076. [MR0143282](#)
- [10] ROSENBLATT, M. (1956). Remarks on some non-parametric estimates of a density function. *Ann. Math. Stat.* **3** 832–837. [MR0079873](#)
- [11] SCHEFF, H. (1947). A useful converge theorem for probability distributions. *Ann. Math. Stat.* **3** 434–438. [MR0021585](#)
- [12] SILVERMAN, B. W. (1986). Nonparametric estimation for statistics and data analysis. Chapman & Hall, New York. [MR0848134](#)
- [13] SILVERSTEIN, J. W. (1995). Strong convergence of the limiting distribution of the eigenvalues of large dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. [MR1370408](#)
- [14] SILVERSTEIN, J. W. and CHOI, S. I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.* **54** 295–309. [MR1345541](#)
- [15] TSE, D. and HANLY, S. (1999). Linear multiuser receivers: Effective interference, effective bandwidth and user capacity. *IEEE Trans. Inform. Theory* **45** 641–657. [MR1677023](#)
- [16] YIN, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20** 50–68. [MR0862241](#)

B.-Y. JING  
 Q.-M. SHAO  
 DEPARTMENT OF MATHEMATICS  
 HONG KONG UNIVERSITY  
 OF SCIENCE AND TECHNOLOGY  
 CLEAR WATER BAY  
 KOWLOON  
 HONG KONG  
 E-MAIL: [majing@ust.hk](mailto:majing@ust.hk)  
[maqms hao@ust.hk](mailto:maqms hao@ust.hk)

G. PAN  
 DIVISION OF MATHEMATICAL SCIENCES  
 SCHOOL OF PHYSICAL  
 AND MATHEMATICAL SCIENCES  
 NANYANG TECHNOLOGICAL UNIVERSITY  
 SINGAPORE 637371  
 E-MAIL: [gmp an@ntu.edu.sg](mailto:gmp an@ntu.edu.sg)

W. ZHOU  
 DEPARTMENT OF STATISTICS  
 AND APPLIED PROBABILITY  
 NATIONAL UNIVERSITY OF SINGAPORE  
 SINGAPORE 117546  
 E-MAIL: [stazw@nus.edu.sg](mailto:stazw@nus.edu.sg)